Longitudinal Linear Models - Some Theory

Terminology

1) Subjects - Clusters
2) Occasions - individual observations within clusters.
3) Levels of variables:
   - Time varying - micro - level 1 - vary by occasion
     e.g. age
   - Time invariant - macro - level 2 - vary by subject
     - "contextual" variables
     e.g. Gender

Simple example

- Time-varying variables
- All variables have random effects.

J subjects measured on T occassions.

Model for j th subject: (Level I Model)

$$\underset{n_j\times 1}{\underset{\sim}{Y}_j} = \underset{n_j\times p}{X_j}\,\underset{p\times 1}{\underset{\sim}{\beta}_j} + \underset{n_j\times 1}{\underset{\sim}{\varepsilon}_j} \qquad \underset{\sim}{\varepsilon}_j \sim N(\underset{\sim}{0}, \sigma I)$$

BLUE of $\underset{\sim}{\beta}_j$ using Y_j and X_j

$$\hat{\underset{\sim}{\beta}}_j = (X_j' X_j)^{-1} X_j' \underset{\sim}{Y}_j$$

Best among Linear Unbiased Estimators repeatedly sampling from subject j

$$E(\hat{\underset{\sim}{\beta}}_j \mid \underset{\sim}{\beta}_j) = \underset{\sim}{\beta}_j$$

$$Var(\hat{\underset{\sim}{\beta}}_j \mid \underset{\sim}{\beta}_j) = \sigma^2 (X_j' X_j)^{-1}$$

Model for $\underset{\sim}{\beta}_j$ (Level 2 Model)

$$\underset{\sim}{\beta}_j \sim N(\underset{\sim}{\gamma}, G)$$

or

$$\underset{\sim}{\beta}_j = \underset{\sim}{\gamma} + \underset{\sim}{\delta}_j \qquad \underset{\sim}{\delta} \sim N(\underset{\sim}{0}, G)$$

So

$$\underset{\sim}{Y}_j = X_j \underset{\sim}{\beta}_j + \underset{\sim}{\varepsilon}_j = X_j(\underset{\sim}{\gamma} + \underset{\sim}{\delta}_j) + \underset{\sim}{\varepsilon}_j$$

$$= \underbrace{X_j \underset{\sim}{\gamma}}_{\text{fixed effects}} + \underbrace{X_j \underset{\sim}{\delta}_j + \underset{\sim}{\varepsilon}_j}_{\text{random effects}}$$

Putting clusters together:

$$\begin{bmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \\ \vdots \\ \underset{\sim}{Y}_j \\ \vdots \\ \underset{\sim}{Y}_J \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \\ X_j \\ \\ X_J \end{bmatrix} \underset{\sim}{\gamma} + ?$$

Putting clusters together:

$$
\begin{bmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \vdots \\ \underset{\sim}{y}_j \\ \vdots \\ \underset{\sim}{y}_J \end{bmatrix}
=
\begin{bmatrix} X_1 \\ X_2 \\ \\ X_j \\ \\ X_J \end{bmatrix} \underset{\sim}{\gamma}
+
\begin{bmatrix} X_1 \underset{\sim}{\delta}_1 \\ X_2 \underset{\sim}{\delta}_2 \\ \vdots \\ \\ X_2 \underset{\sim}{\delta}_j \end{bmatrix}
+
\begin{bmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \\ \vdots \\ \\ \underset{\sim}{\varepsilon}_J \end{bmatrix}
$$

Putting clusters together:

$$\begin{bmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \\ \vdots \\ \underset{\sim}{Y}_j \\ \vdots \\ \underset{\sim}{Y}_J \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \\ X_j \\ \\ X_J \end{bmatrix} \underset{\sim}{\gamma} + \begin{bmatrix} Z_1 & O & O & \cdots & O \\ O & Z_2 & O & & O \\ O & O & \ddots & & \\ \vdots & & & Z_j & \\ \vdots & & & & \ddots \\ O & & & & Z_J \end{bmatrix} \begin{bmatrix} \underset{\sim}{\delta}_1 \\ \delta_2 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_J \end{bmatrix} \quad \begin{bmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \\ \vdots \\ \underset{\sim}{\varepsilon}_J \end{bmatrix}$$

Putting clusters together:

$$\begin{bmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \\ \vdots \\ \underset{\sim}{Y}_j \\ \vdots \\ \underset{\sim}{Y}_J \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \\ X_j \\ \\ X_J \end{bmatrix} \underset{\sim}{\gamma} + \begin{bmatrix} Z_1 & 0 & 0 & \cdots & 0 \\ 0 & Z_2 & 0 & & 0 \\ 0 & 0 & \ddots & & \\ \vdots & & & Z_j & \\ & & & & \ddots \\ 0 & & & & Z_J \end{bmatrix} \begin{bmatrix} \underset{\sim}{\delta}_1 \\ \underset{\sim}{\delta}_2 \\ \vdots \\ \underset{\sim}{\delta}_j \\ \vdots \\ \underset{\sim}{\delta}_J \end{bmatrix} \quad \begin{bmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \\ \vdots \\ \underset{\sim}{\varepsilon}_J \end{bmatrix}$$

$$\boxed{\underset{\sim}{Y} = X\underset{\sim}{\gamma} + Z\underset{\sim}{\delta} + \underset{\sim}{\varepsilon}}$$

We use Z_i instead of X_i to anticipate that Z_i may differ from X_i.

Often J is large and γ, G, σ^2 are estimated with much more precision than $\hat{\beta}_j$.

Pretend for this discussion that γ, G, σ^2 are "known".

Can we improve on $\hat{\underset{\sim}{\beta}}_j$ as an estimator of $\underset{\sim}{\beta}_j$?

What do we know about $\underset{\sim}{\beta}_j$?

1) $E(\hat{\underset{\sim}{\beta}}_j - \underset{\sim}{\beta}_j) = \underset{\sim}{0}$ with $Var(\hat{\underset{\sim}{\beta}}_j - \underset{\sim}{\beta}_j) = \sigma^2(X_j'X_j)^{-1}$

2) $\underset{\sim}{\beta}_j$ comes from a population with mean $\underset{\sim}{\gamma}$ and variance G.

So $E(\underset{\sim}{\gamma} - \underset{\sim}{\beta}_j) = \underset{\sim}{0}$ with $Var(\underset{\sim}{\gamma} - \underset{\sim}{\beta}_j) = G$

So $\hat{\underset{\sim}{\beta}}_j$ & $\underset{\sim}{\gamma}$ are unbiased predictors of $\underset{\sim}{\beta}_j$ (treating $\underset{\sim}{\beta}_j$ as random

How to combine them?

Best Linear Unbiased Linear Combination Predictor

BLUP: Take a weighted average with weights proportional to "precision" $= (\text{Variance})^{-1}$

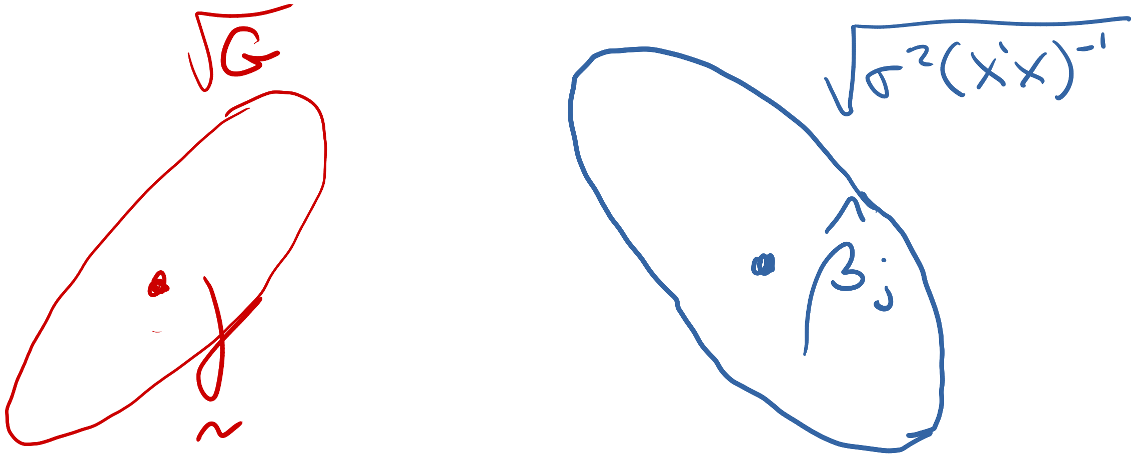
BLUP

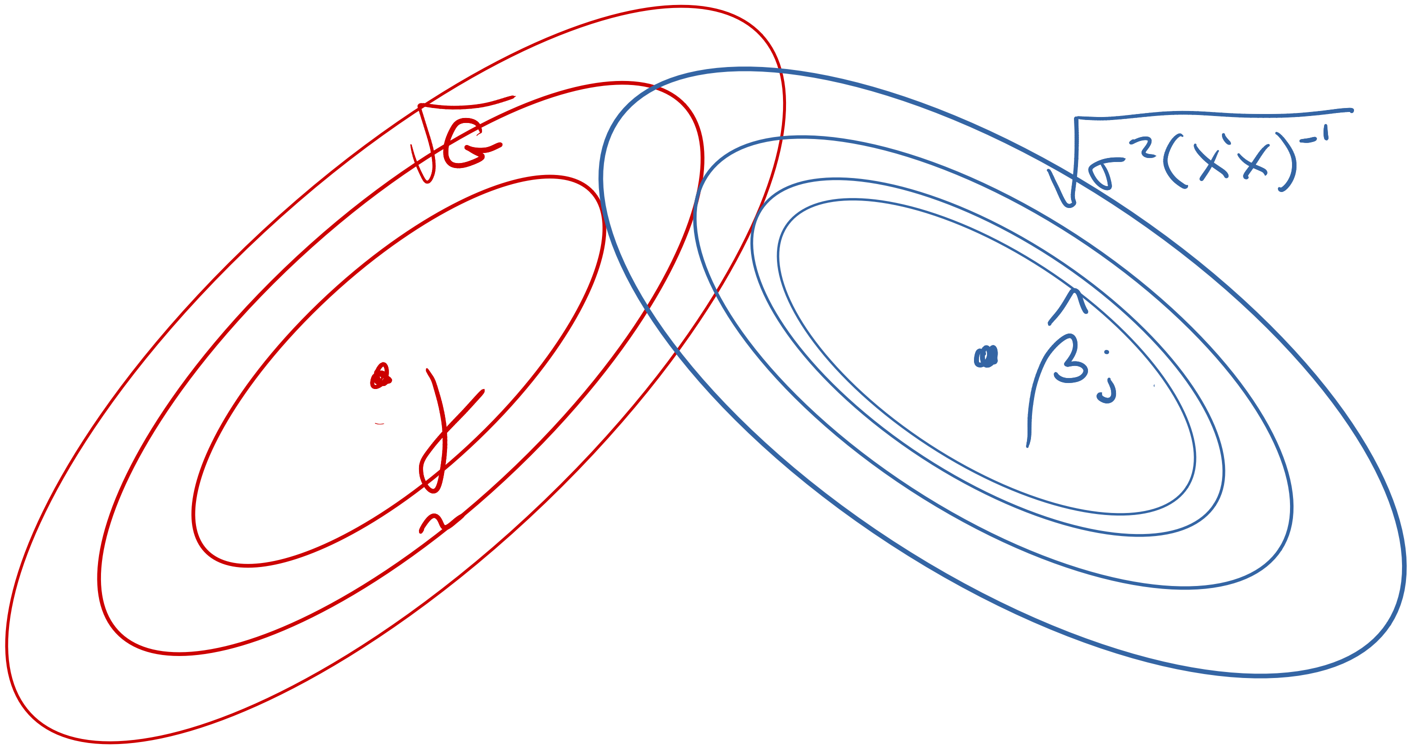
$$\tilde{\beta}_j = (W_1 + W_2)^{-1}(W_1 \tilde{y} + W_2 \hat{\beta}_j)$$

$$W_1 = G^{-1} \qquad W_2 = [\sigma^2 (X_j' X_j)^{-1}]^2$$

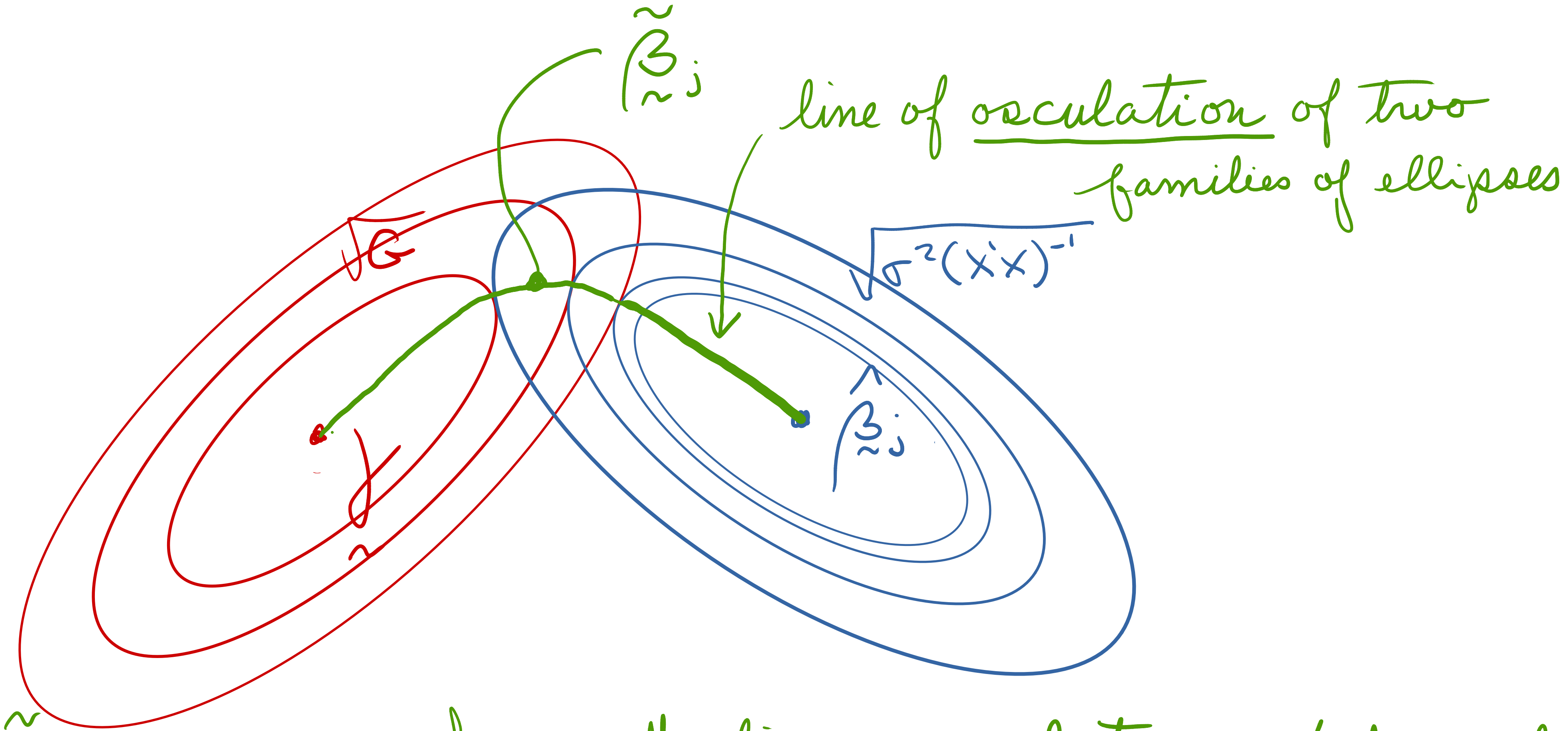
$$= \frac{1}{\sigma^2} X_j' X_j$$

What does this look like?



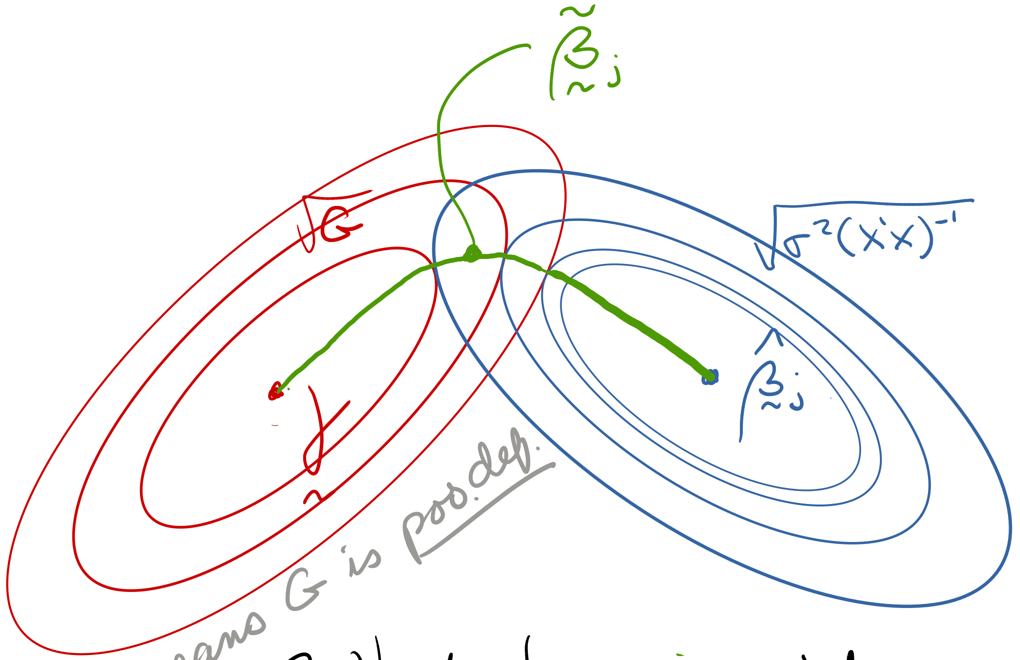


$\sqrt{G}$
$\tilde{\gamma}$
$\hat{\beta}_j$
$\sqrt{\sigma^2(X'X)^{-1}}$



$\tilde{\beta}_j$ will be somewhere on the line of osculation of the families

$\gamma \oplus k\sqrt{G}$ and $\beta \oplus k\sqrt{\sigma^2(X_j'X)^{-1}}$



- This is often referred to as "shrinking" $\hat{\beta}_j$ towards γ
- $\tilde{\beta}_j$ is a type of "shrinkage" estimator

This means G is pos. def.

What happens if:

$n_j \to \infty$ and $G > 0$ } $X'X \to n\Pi_X$ (asymptotic inner product matrix) so $\frac{1}{\sqrt{n}}\sqrt{\sigma^2\Pi} \to 0$ and $\tilde{\beta}_j \to \hat{\beta}_j$

G nearly singular with one small radius } $\tilde{\beta}_j$ will be close to the space $\perp$ small radius.

n_j is small } $\tilde{\beta}_j$ will be dominated by γ

Q